Introduction to KISTI National Supercomputing Infrastructure

Taeyoung Hong

Supercomputing Infrastructure Center,
Division of National Supercomputing, KISTI

KiSTi
www.kisti.re.kr
한국과학기술정보연구원
Korea Institute of Science and Technology Information

KISTI
SINCE 1962
60

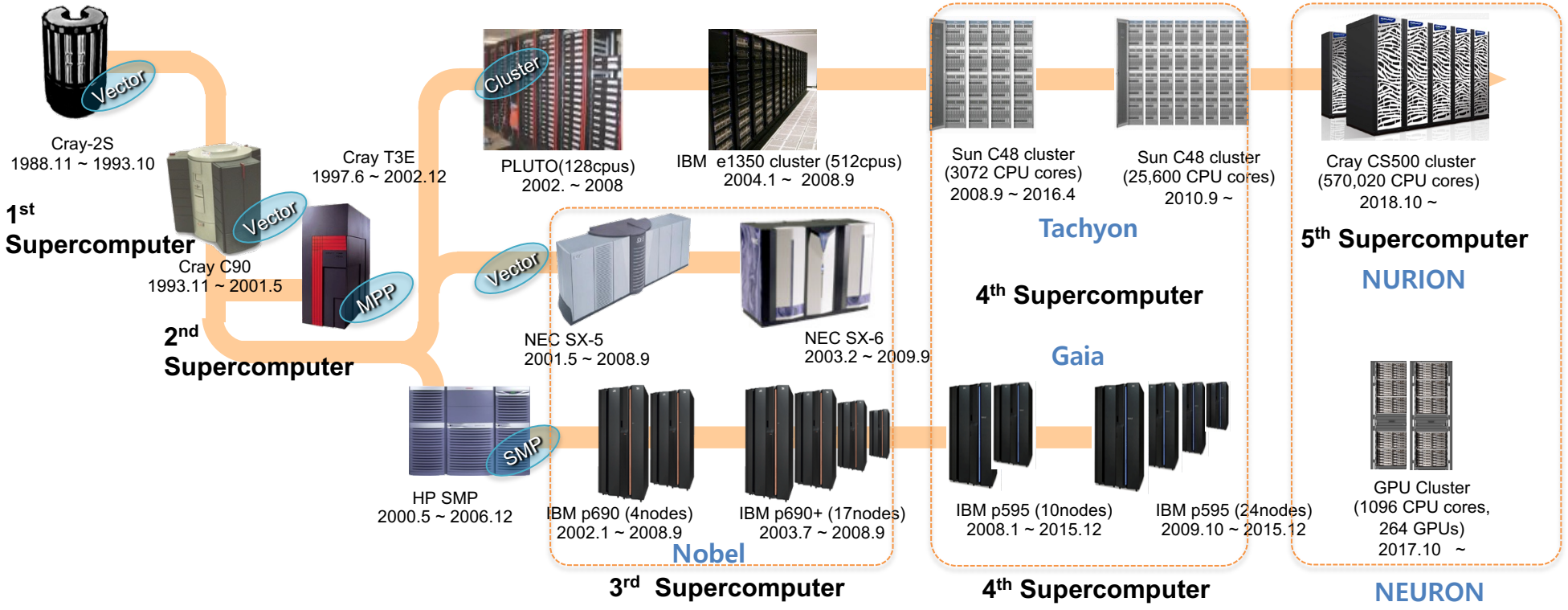
KiSTi 한국과학기술정보연구원
Korea Institute of Science and Technology Information
www.kisti.re.kr

KISTI SINCE 1962

CONTENTS

KISTI National Supercomputing Center

- KISTI has served a national supercomputing center of S. Korea, providing both high performance computing and storage resources to Korean researchers and engineers since 1988.
- Starting with Cray-2S, Korea's first supercomputer, HPC system have been continuously upgraded on the regular basis. Currently, we are operating our 5th supercomputer - **NURION**
- **NURION** was ranked at the 11th in Top500 (2018.11) and its rank now declined to 61th (2023.11).

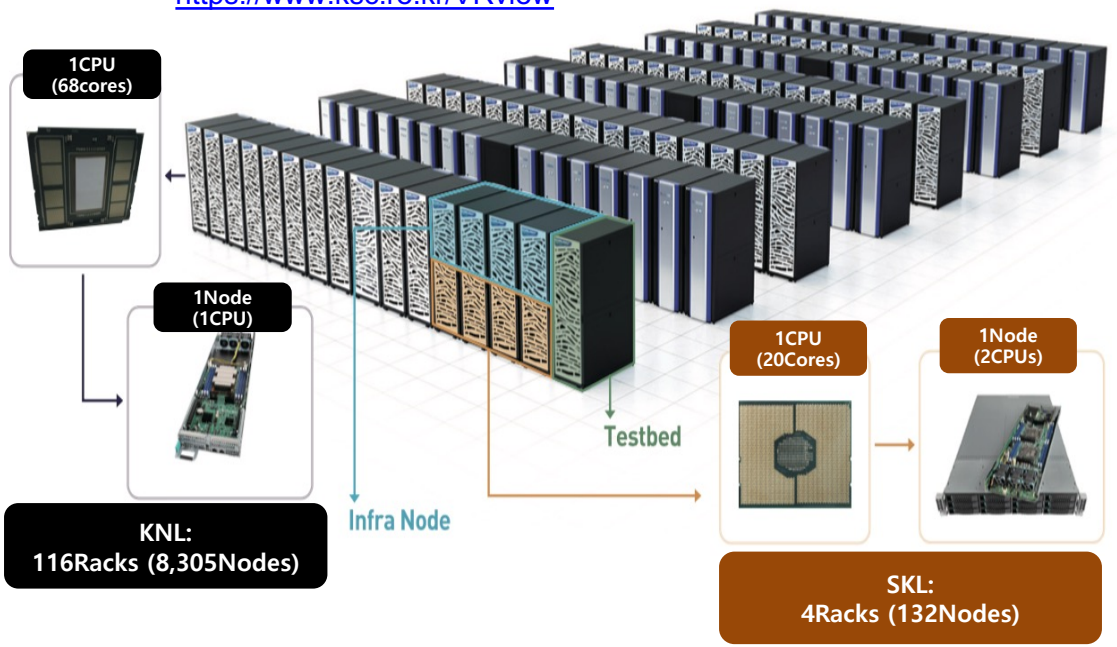


KISTI-5 Supercomputer NURION

- In Dec. 2018 **NURION** system started to put in production.
- It provides **25.7 PFlops** computing power, **33.88 PB** storage capacity with **12.3 GB/s** interconnect network bandwidth

- 128 Racks of Computing Components
  - 8 rows, with 16 Cabinets in each
  - 8,305 KNL Compute Nodes
  - 132 Xeon Skylake CPU Nodes
- 12 Racks of DDN Storage
  - 20 PB of Scratch Storage
  - 1 PB of Home and App Storage
  - 0.8 PB Burst Buffer
- TS-4500 Tape Library
  - 10PB / 1,700 Media
- Interconnection Network
  - Intel Omni-Path Architecture (100Gbps)

Click to view **NURION**
https://www.ksc.re.kr/VRview



KISTI-5 Specification

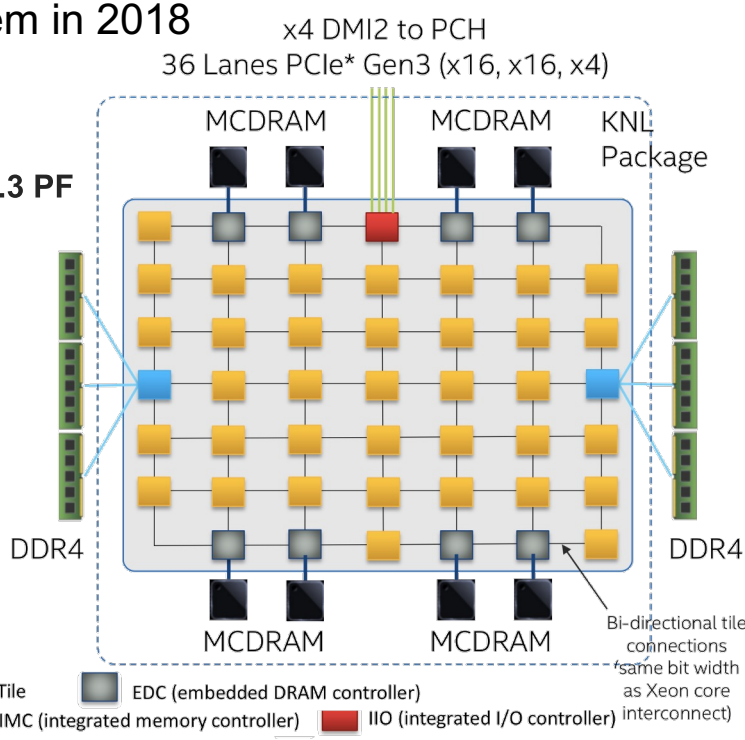
KISTI SINCE 1962

- One of the Largest (KNL/OPA based) Off-the-shelf Cluster System in 2018
- Rpeak 25.7PFlops, Rmax 13.9PFlops

Compute nodes

8,305 Many-core CPU Nodes, 116 Racks, 25.3 PF

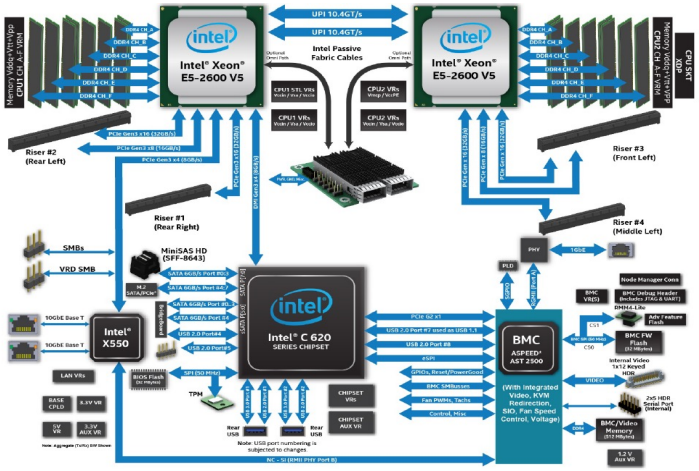
- 1x Intel Xeon Phi KNL 7250, 68Cores,1.4GHz
- 3TFlops Peak, ~0.2 Bytes/Flops Memory BW
- 96GB (6x16GB) DDR4-2400 6 channel RAM
- 16GB MCDRAM(HBM, 460GB/s)
- 1x 100Gbps OPA HFI, 1x On-board GigE Port



CPU-only nodes

132 Skylake CPU Nodes, 4 Racks, 0.4 PF

- 2x Intel Xeon SKX 6148, 20Cores, 2.4GHz
- 192GB (12x 16GB) DDR4-2666 RAM
- 1x Single-port 100Gbps OPA HFI card
- 1x On-board GigE (RJ45) port



Manycore CPU (KNL)
& Commodity Server
CPUs
↓
(KISTI-6)
HPC GPUs
& Commodity Server
CPUs

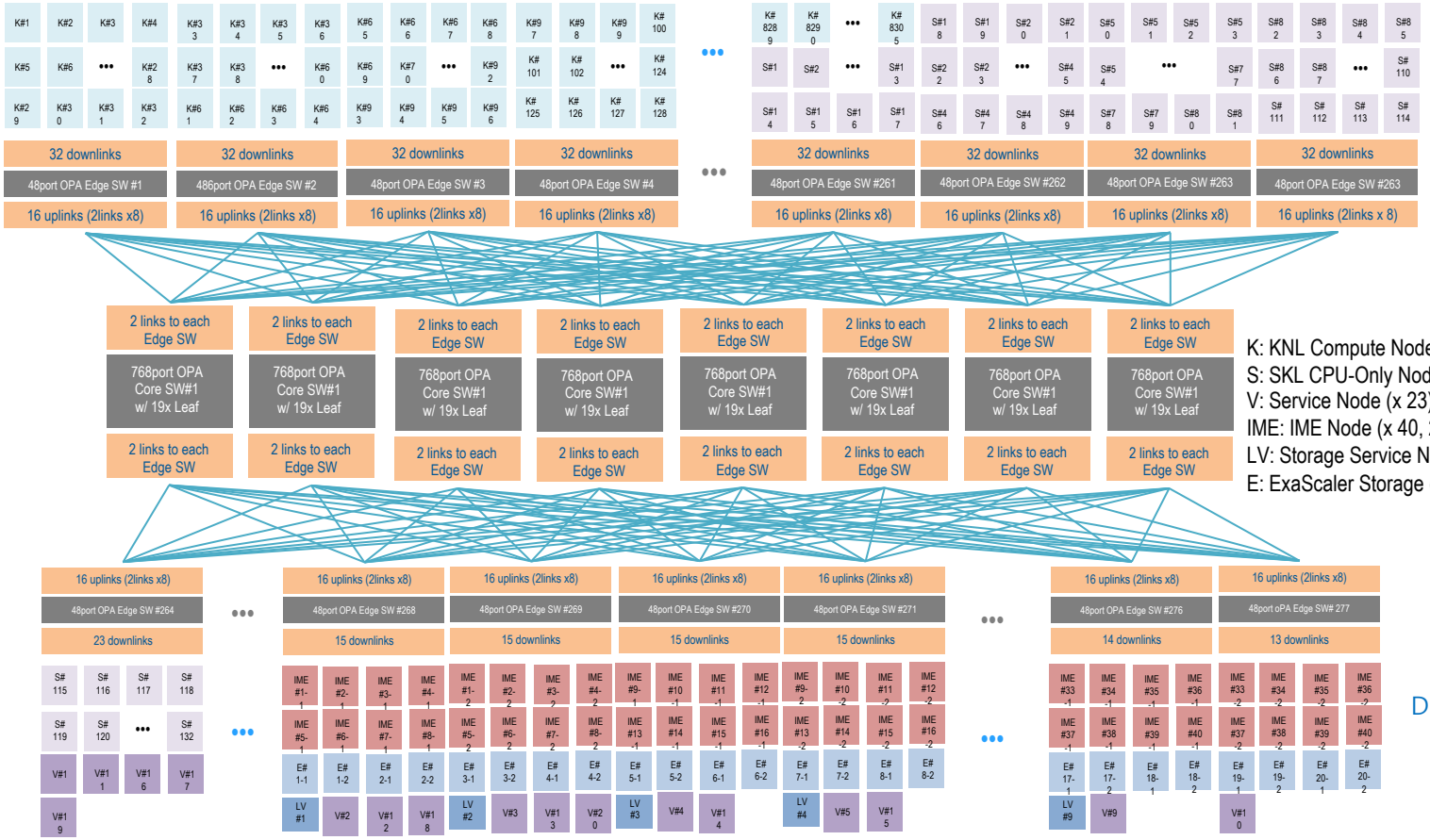
KISTI-5 Specification

KISTI SINCE 1962

Interconnect

OPA@12.3GB/s, Fat-Tree, 50% Blocking

- Intel OPA interconnect: 274x 48-port OPA edge switches, 8 x 768-port OPA core switches
- Bandwidth: 12.3 GB/s, Bisectional Bandwidth: 27 TB/s
- 10^{-16} BER (Bit Error Rate), Adaptive routing



K: KNL Compute Node (x 8305)
S: SKL CPU-Only Node (x 132)
V: Service Node (x 23)
IME: IME Node (x 40, 2x connections/each node)
LV: Storage Service Node (x 9)
E: ExaScaler Storage (x 40)

Fat-tree Topology
IB/OPA
↓
(KISTI-6)
Dragonfly(+) Topology,
Slingshot, IB NDR

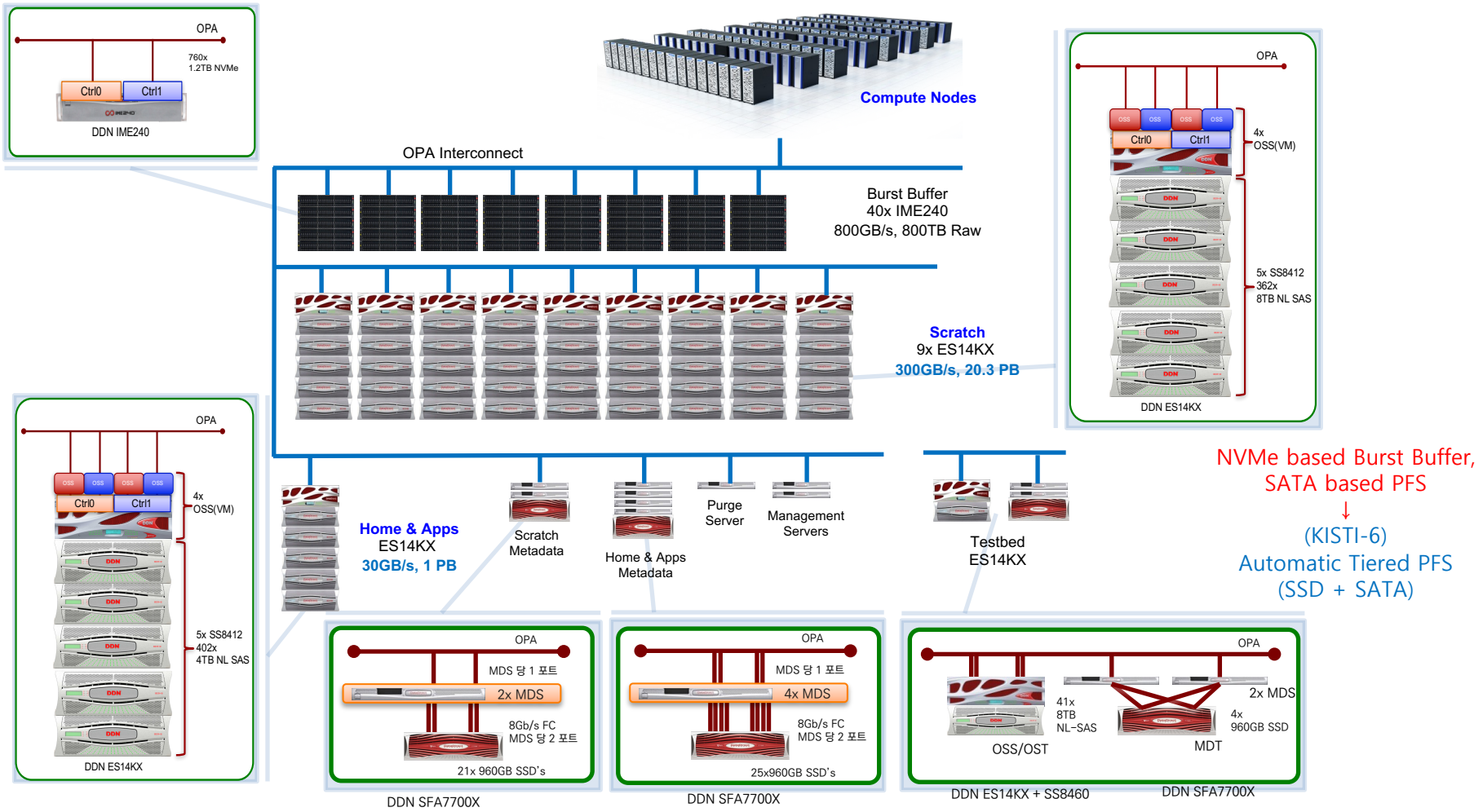
KISTI-5 Specification

KISTI SINCE 1962

Storage

20PB PFS@300GB/s, 0.8PB Burst Buffer@800GB/s, 10PB Archiving@10GB/s

- ☑ Global Scratch: 20PB, 300GB/s, Home and Application : 1PB (Scratch is one of the largest single shared PFS in 2018)
- ☑ NVMe Burst Buffer: 0.8PB, 800GB/s, Cray TSMSF and IBM TS4500 (10PB) (Storage N/W is non-blocking)



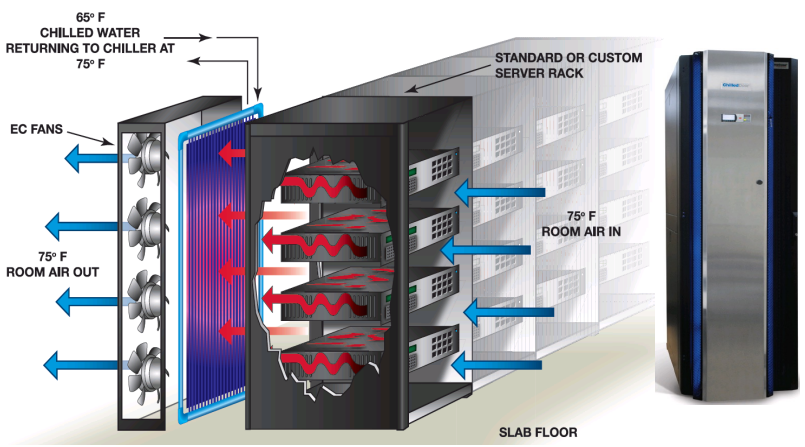
KISTI-5 Power Supply and Cooling

Power Supply Equipment

- 20MW dedicated dual power lines from 2 different power substations to KISTI facility
- 11.4MW UPS for at least 20 minutes of backup power for computing/storage system
- 5MVA diesel power generators for cooling and power distribution units during power outages

Mechanical Cooling Equipment

- Chilled water-cooled systems with rear doors
- PUE(Power Usage Effectiveness) reduced from 1.58 (KISTI-4) to 1.34 (KISTI-5) by applying economizer(water side), free cooling chiller, and dynamic power management of chillers depending on system load



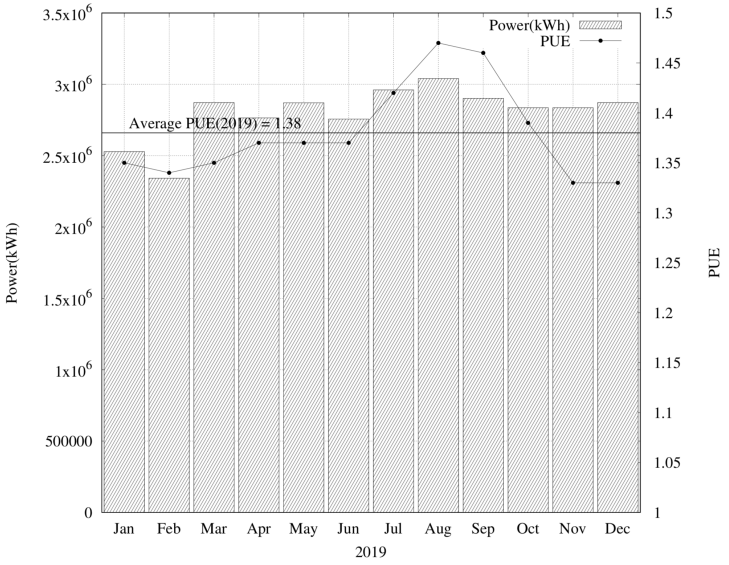


Fig. 3: Energy consumption and PUE of the Nurion system in 2019.

UPS for whole system
↓
(KISTI-6)
UPS only for
storage

Indirect water
Cooling &
(PUE 1.3)
↓
(KISTI-6)
Direct water
Cooling
(PUE 1.1)

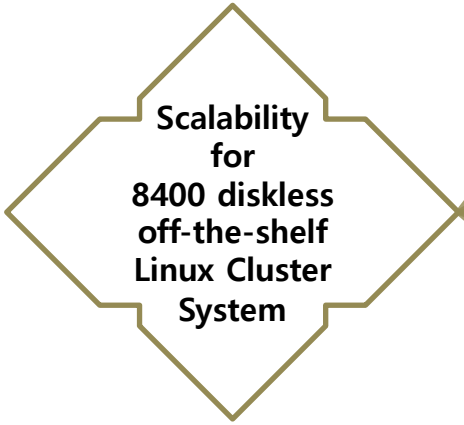
KiSTi 한국과학기술정보연구원
Korea Institute of Science and Technology Information
www.kisti.re.kr

KISTI SINCE 1962

CONTENTS

System Software Optimization

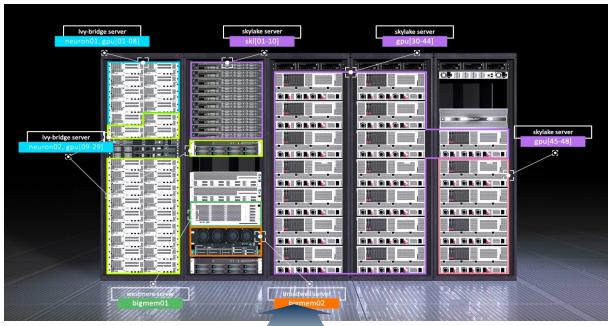
- For speed-up of provisioning OSes into diskless compute nodes, 132 Skylake nodes are configured as relay servers used in rebooting on periodic maintenance window
- The ethernet network of the whole system is split into 16 VLANs to avoid MAC flooding due to the limit of the MAC entries of ethernet switches
- The delay of start-up and run time of MPI jobs due to PBS integration with MPI has been enhanced over time

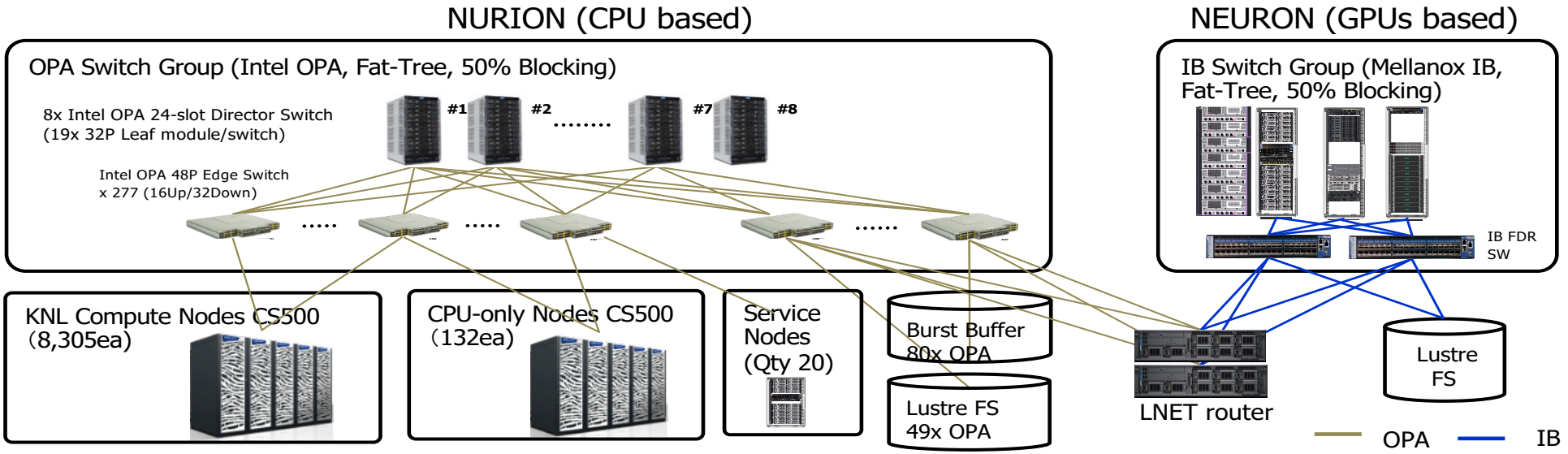


Specification	Products
Cluster Manager (Provisioing, Mgmt, OpenStack)	Bright Computing
Operating System	CentOS
Workload Manager	PBS Pro
Compilers	Intel, Cray PE, GNU
MPI Libraries	Intel MPI, Open MPI, MVAPICH2
Interconnect Software	Intel OPA
Parallel File System	Lustre
Burst Buffer	IME
Debugger	Allinea DDT
Profiler	CrayPat, Intel Vtune
User Account Management	LDAP
Commercial Software	ABAQUS, ANSYS, Gaussian
User Development Environment	Module, Conda, Docker/Singularity, MyKSC

GPU-based KISTI-5 Auxiliary System, *Neuron*

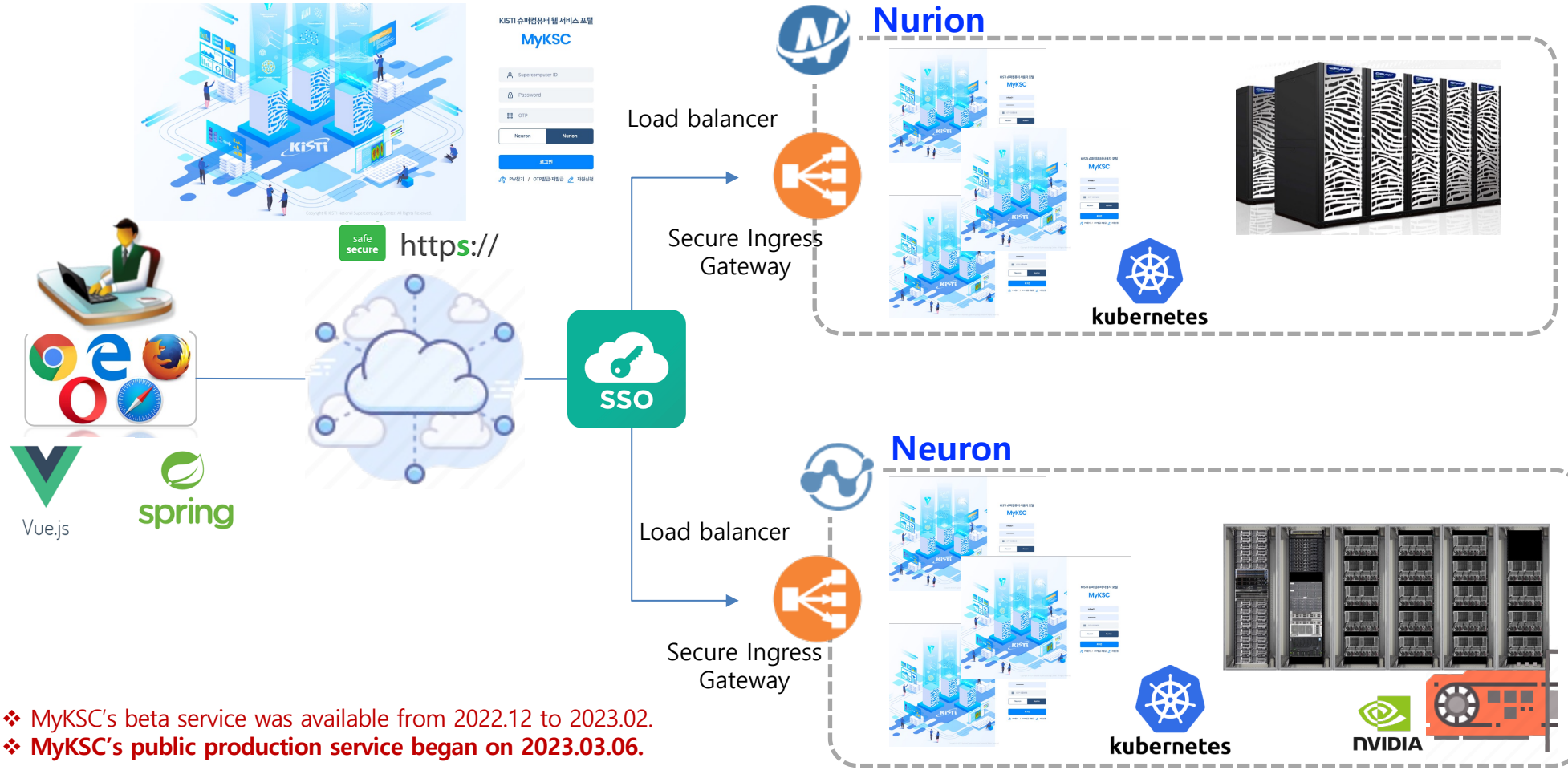
- ☑ Based on GPU and next-generation hardware (official service in July 2019)
  - 65 servers, **260 GPUs (140 V100, 120 A100),** total **3.53PFlops**
  - **+36 Nvidia H100 or AMD MI300 GPUs** will be added **in 2024**
- ☑ Focused on AI/DL R&D and MD/DFT fields showing good performance acceleration from GPUs
- ☑ Filesystem sharing support with NUIRON Lustre PFS
- ☑ Supported parallel AI execution (Horovod, Pytorch DDP, TF Distributed …)
- ☑ Supported other development environment (Conda, Docker/Singularity, Module …)
- ☑ **To be expanded to +10 PFlops (FP64) by annual upgrade until 2025**





MyKSC: KISTI Supercomputer Web-based Service Portal

- To provide users with easy and secure access to a range of development tools and services for HPC and ML/DL, Big data analysis
- Supports Web terminal (WeTTY), Data Management (File Run), Jupyter Lab, RStudio, VS code, Remote Desktop (VNC), Batch Job Launch & Management(PBS/Slurm)
- SSO (Single-Sign On), Load Balancer, Secure Ingress Gateway, and Kubernetes



- MyKSC's beta service was available from 2022.12 to 2023.02.
- **MyKSC's public production service began on 2023.03.06.**

KiSTi 한국과학기술정보연구원
Korea Institute of Science and Technology Information
www.kisti.re.kr

KISTI SINCE 1962

CONTENTS

Applications for Resource Allocation

Open call based User Support Program, 'R&D innovation Support Projects' (Free of Charge, 90% of Resources)

How to apply for R&D Innovation Support Projects

The call for applications will be posted on this website. Submit your proposals for the program application on the page.

· Biannual Program Sections

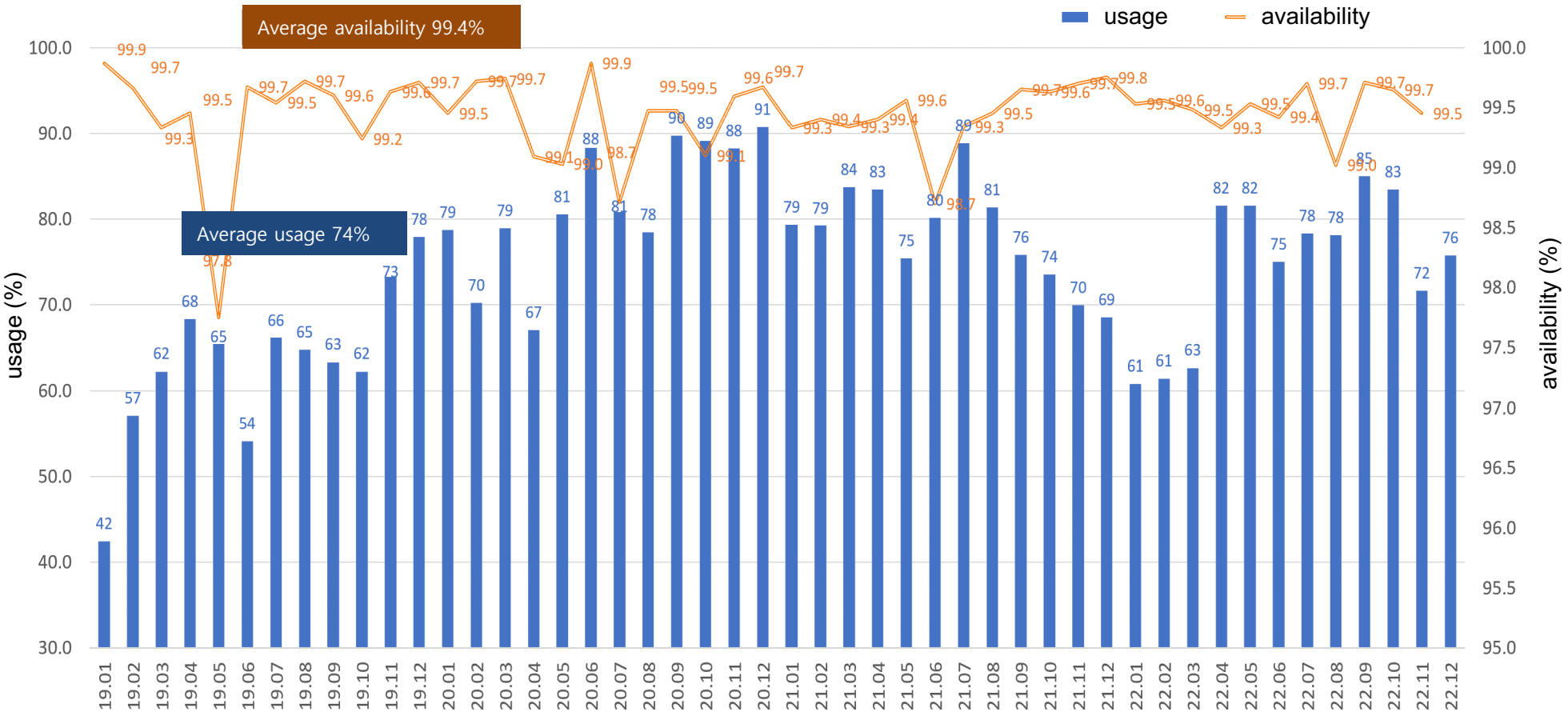
	Grand Challenge	Creative research
Target	Researchers and associations need 5PF or higher computing power to solve massive scientific problems	performing R&D projects in KISTI supercomputers
Requirements	Computational problems have to used 5PF or larger at once time	Doctoral degree and Korean nationality are required
Resources/period	Dedicated resources and flexible support in three months at maximum	Shared resources in one year, no additional resources can be added

Annual Subscription based paid support Program (Pre-paid, Production-cost level price policy, 10% of Resources)

- Researchers or institutions need Resources without going through expert evaluation to protect core research contents or corporate te chnology
- Researchers or institutions want to use computational resources exclusively for a certain period of time for the nature of their researc h such as mission-critical applications

KISTI-5 Operation and Service

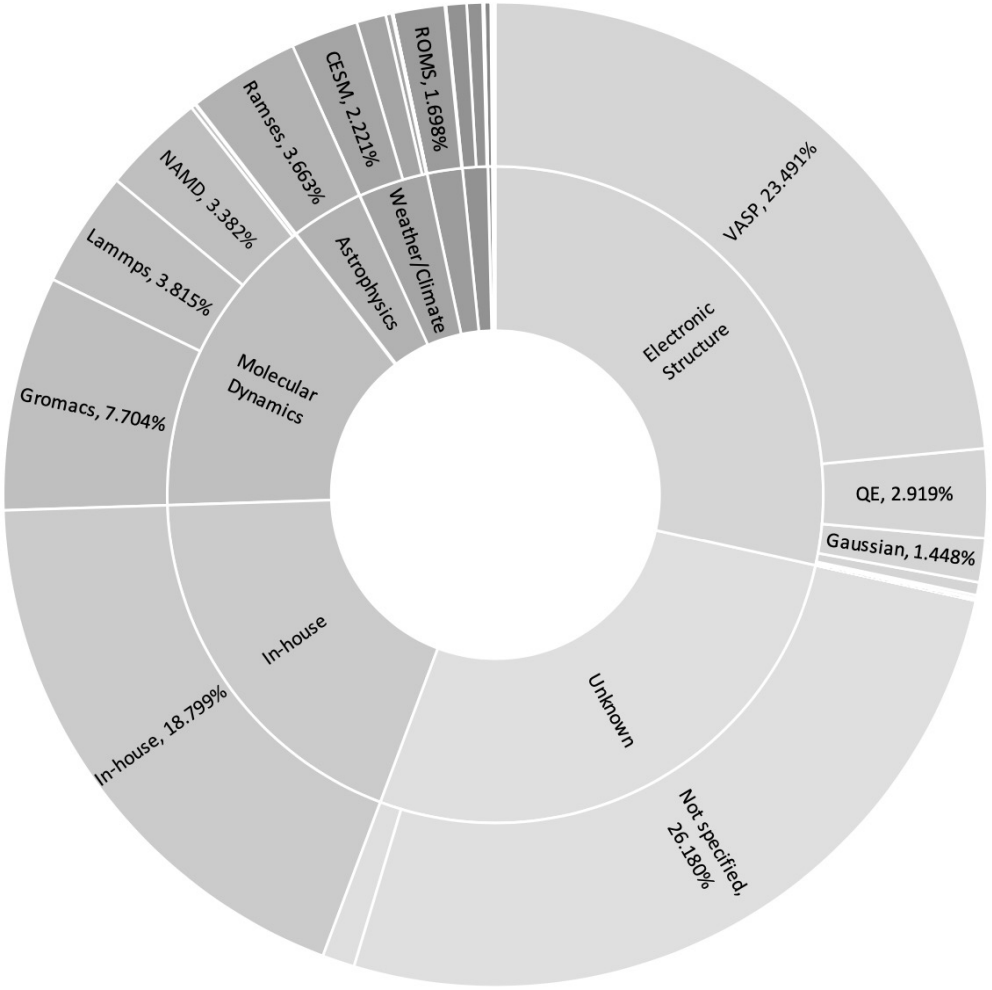
- Average availability 99.4%, average usage rate 74% (As of 2019.01~ 2023.12)
- From Dec. 2018 up to now 11 million user jobs have executed



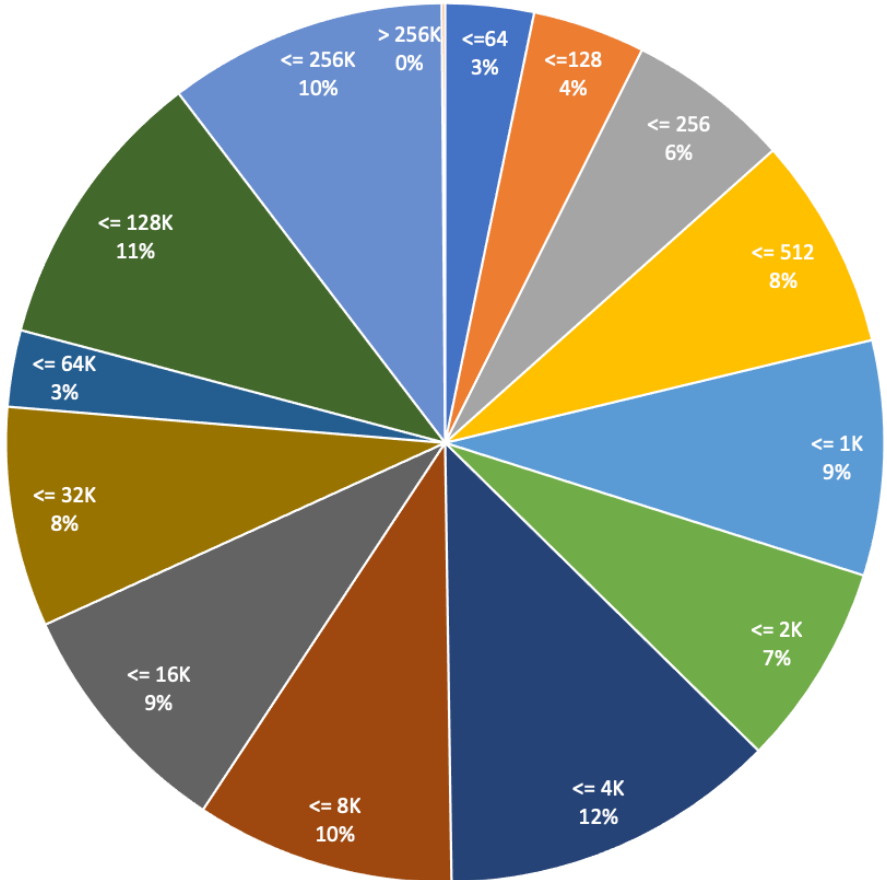
Nurion System Production Status

- +300M node-hours provided and +11M user jobs have run from Dec 2018 to now.
- Average job size of all user jobs in 2019 is 631CPU cores(arithmetic average), 699(jobs' wall-time weighted average) and 34,313(jobs' CPUs x wall-time weighted average)

KISTI-5 current workload

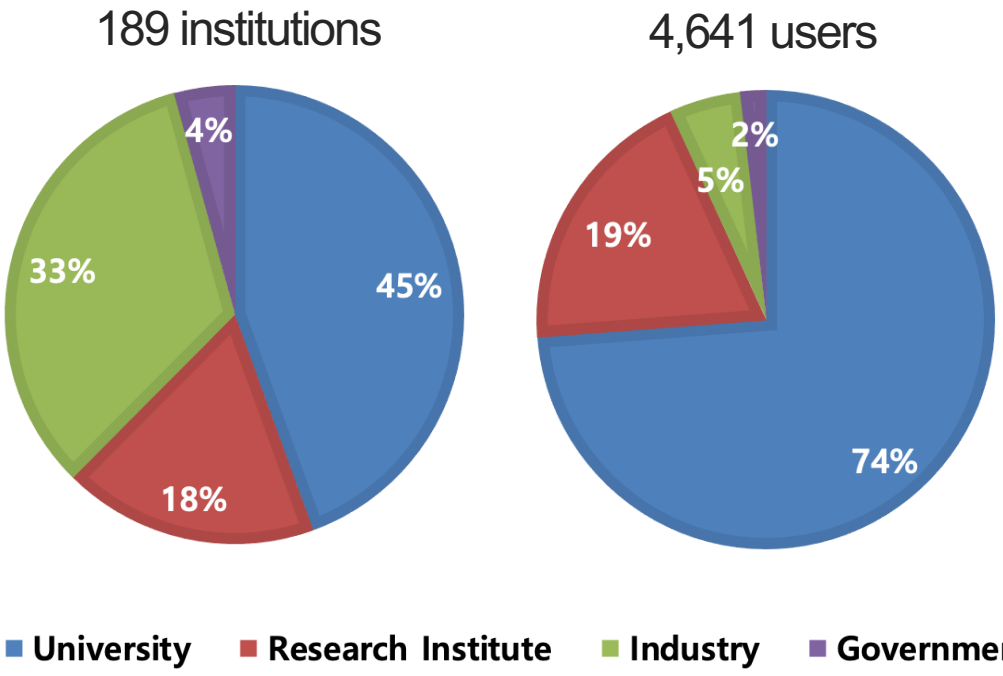


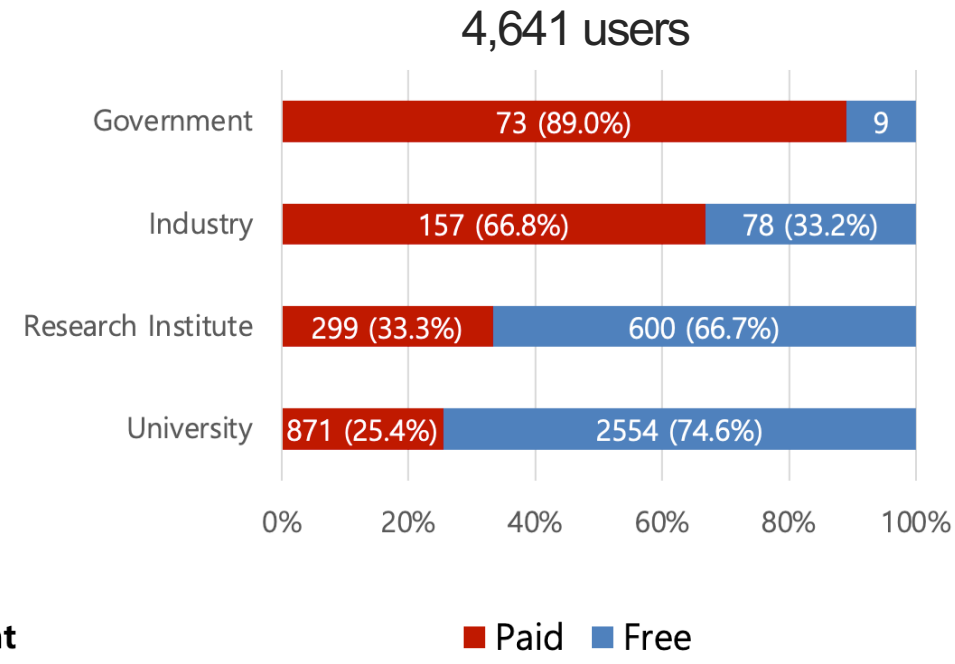
User job distribution categorized by user job size (2019)



User Distribution

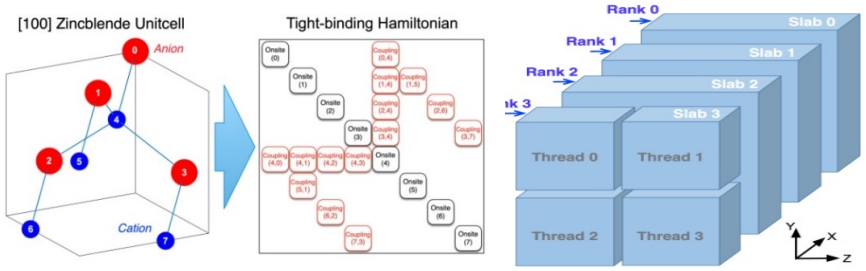
- 189 institutions and 4,641 users nationwide have used KISTI-5 and Neuron in 2022
- Supported HPC resources and services to various institutions such as universities, research institutes, industries, and government agencies nationwide.





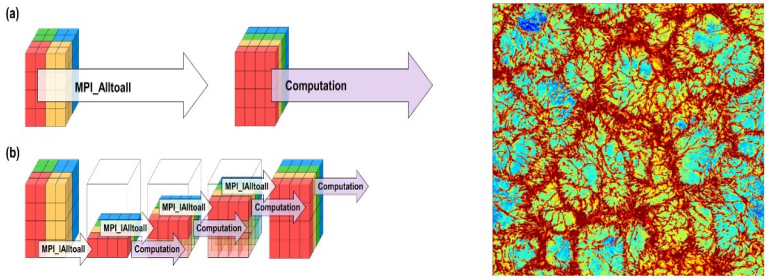
User Support and R&D Collaboration

The world largest scale simulation of atomic structure of semiconductor



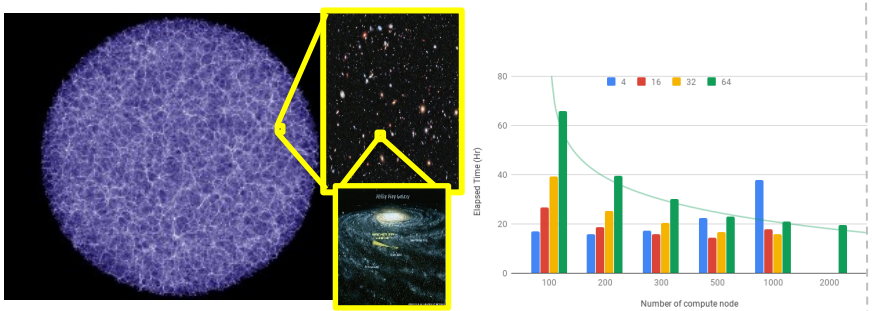
- Matrixization and 3D domain decomposition
- Electronic structure of a silicon box consisting of 400 million atoms simulated (the previous record: 51million atoms)

The world class parallelization of fluid analysis of turbulence



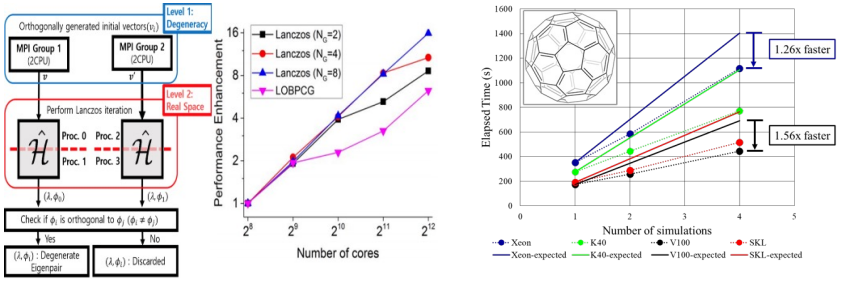
- Optimization/parallelization of CFD in-house code (DNL-TBL) to scale to 2,500 compute nodes

The parallelization of Astrophysics code(RAMSES) with 7.5PFlops



- Break-through the limited scalability of cosmological numerical simulation (RAMSES)
- World largest scale(1kpcs) cosmology simulation by using 2,500 nodes (7.5PF)

Joint development of quantum chemistry simulation software



- Joint development of Quantum Chemistry software and it's optimization and parallelization
- The development of AI based electronic structure code

KiSTi 한국과학기술정보연구원
Korea Institute of Science and Technology Information
www.kisti.re.kr

KISTI SINCE 1962

CONTENTS

KISTI-6 National Flagship Supercomputer

KISTI SINCE 1962

- A system with **600PF (FP64 matrix)** that can simultaneously respond to huge computational demands of HPC and AI communities
- The beta service will be started **in 2025**

Computational

Performance : 600PFlops (FP64)

- **CPU-only Partition**
- **GPU Partition**
- **Storage (Flash 10PB, HDD 190PB)**
- **Interconnect Network (400Gbps)**

Hetero computing environment equipped with CPUs and GPUs

- By setting the **CPU:GPU number ratio to 1:2,** the performance target and sufficient number of CPUs are secured and ensure compatibility of existing user code.
- Estimation of storage capacity for large-scale HPC and AI (**+200PB**)
- Ensure balanced performance of compute and storage networks and reflect the latest interconnect roadmap (**Dragonfly(+) or Fat-tree**, **400Gbps**, 2 Ports per node)

Facilities

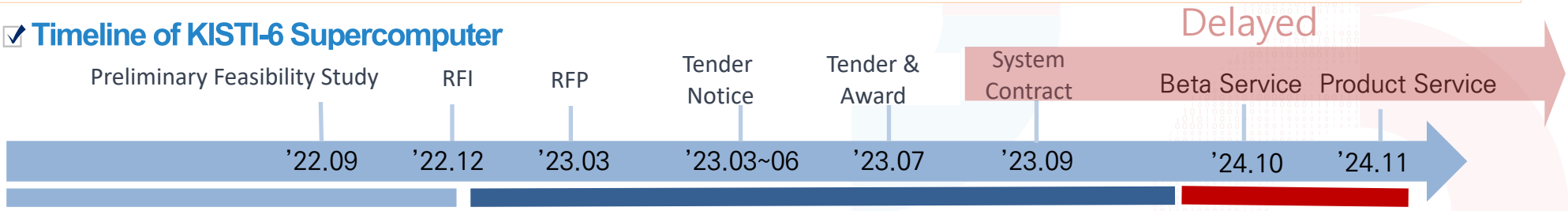
Power and Cooling

- **Power (18MW)**
- **Cooling Capacity (4,000RT)**
- **Space (300m²)**
- **Direct Liquid Cooling**

Providing stable power and cooling with 18MW capacity

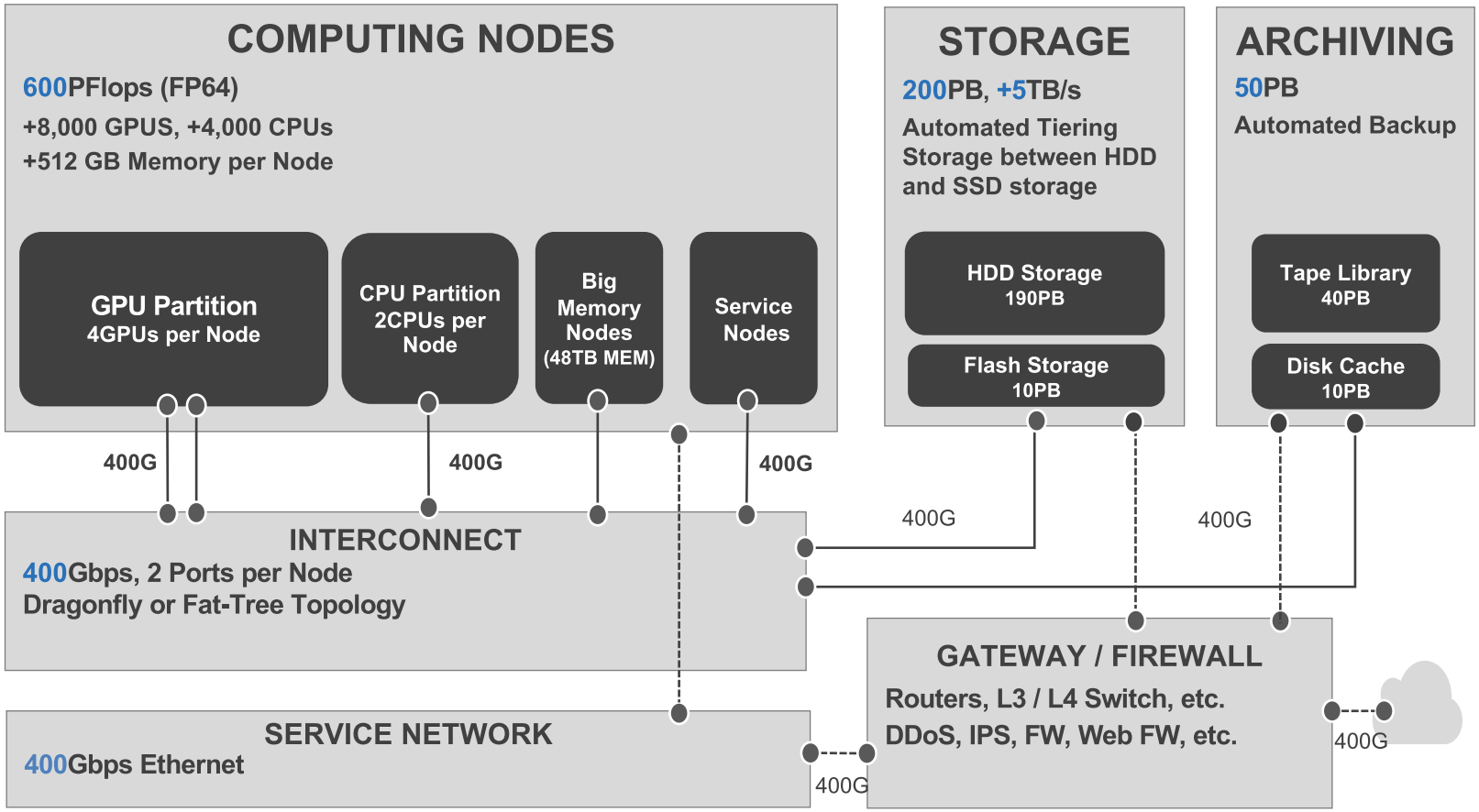
- Reduce construction period and cost by **Utilizing the existing facilities**
- Closed circuit cooling towers will be mainly used for heat-exchange instead of chillers

Timeline of KISTI-6 Supercomputer



KISTI SINCE 1962

System Diagram



KISTI-6 National Flagship Supercomputer Expected Impacts

KISTI SINCE 1962

Efficiency improvement and overcoming limitations in R&D based on HPC, accelerating the acquisition of strategic technology and promoting scientific and technological innovation capacity

Core infrastructure for pioneering future innovative technologies

Support for Strategic Technologies

Providing infrastructure for the development of source technologies such as Quantum, Bio, Semiconductor, AI, etc.

- **Quantum**: Overcoming the limitations of existing Quantum Computing including Quantum emulators on HPC
- **Bio**: Bio-digital convergence including brain science and synthetic biology research, AI-driven drug development
- **Material**: Simulation research based on HPC for securing core tech, such as next-gen semiconductor devices

Use Case Examples

Classification	Research Topics
Bio	AI-based high-speed large-scale genomic analysis
Auto Driving	Accident-prone area safety driving control model and autonomous driving model in complex environments
ICT & AI	NLP, Vision, Multi-modal
Weather/ Climate	Improvement of GPU-based weather/climate models and parameterization of clouds
Manufacturing	large-scale CFD analysis and multi-objective, multi-disciplinary transient optimal design

(Resource Allocation) **National Strategic R&D Sector 50%, Public/Social Sector 10%, Industries 20%, National Shared Utilization 20%**

(Application Area) **Computational Science 70%, AI 30%**

Reducing both the time and cost of R&D through more sophisticated and faster simulations & AI by utilizing HPC

QnA